

Sample PDF Document

This is a sample PDF document created for testing the RAG CLI document loader. The document loader should be able to extract this text from the PDF file.

About RAG Systems

RAG (Retrieval-Augmented Generation) is an AI framework that enhances large language models by providing them with relevant context from a document database. This allows the model to generate more accurate and factual responses.

Key Components

1. Document Ingestion: Process and store documents in various formats
2. Embedding Generation: Convert text into vector representations
3. Vector Storage: Store embeddings in a database like ChromaDB
4. Retrieval: Find relevant documents based on query similarity
5. Generation: Use LLM to generate answers with retrieved context

This document contains test content across multiple lines to verify that the PDF loader correctly extracts and preserves the text structure.

Page 2: Additional Content

This is the second page of the sample PDF document.

Technical Specifications

The RAG CLI system supports the following document formats:

- PDF files (.pdf)
- Markdown files (.md, .markdown)
- HTML files (.html, .htm)

Each document is processed to extract plain text content along with metadata such as title, author, page count, and word count.

The extracted text is then chunked into smaller segments for embedding generation and storage in the vector database.